ADA036842

HOW TO USE MULTI-ATTRIBUTE UTILITY MEASUREMENT FOR SOCIAL DECISION-MAKING

SOCIAL SCIENCE RESEARCH INSTITUTE
UNIVERSITY OF SOUTHERN CALIFORNIA

Ward Edwards

12

See 1473

ADVANCED DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research – Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
The Oregon Research Institute
Perceptronics, Incorporated
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

LT COL Roy M. Gulick, USMC
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

(12)

TECHNICAL REPORT [redacted]
76-3-1-T

HOW TO USE MULTI-ATTRIBUTE UTILITY MEASUREMENT FOR SOCIAL DECISION-MAKING

by

Ward Edwards

Sponsored by

Defense Advanced Research Projects Agency
ARPA Order No. 3052

August 1976

SOCIAL SCIENCE RESEARCH INSTITUTE
University of Southern California
Los Angeles, California 90007

SUMMARY

Decisions do, and should, depend on values and probabilities--both subjective quantities. Public decisions, even more than other kinds, also should depend on values and probabilities. These quantities should be public, not only in the sense of being publishable, but also in the sense that the values, and perhaps the probabilities, that lie behind the decision should depend on some kind of social consensus, or at least on some kind of aggregation of individual views, rather than on any single individual's views.

The thrust of this paper is that a public value is a value assigned to an outcome by a public, usually by means of some public institution that does the evaluating. This amounts to treating "a public" as a sort of organism whose values can be elicited by some appropriate adaptation of the methods already in use to elicit individual values. From this point of view, the interest of the problem lies in finding the appropriate adaptation of those methods, an adaptation that will take into account individual disagreements about values, individual differences in relevant expertise, existing social structures for making public decisions, and problems of feasibility.

Arguments over public policy typically turn out to hinge on disagreements about values. Such disagreements are often about degree, not kind; developed and developing nations may agree on the virtues both of increased industrialization and decreased degradation of the environment, but may differ about the relative importance of these goals. Normally, such disagreements are fought out in the context of specific decisions, over and over again, at enormous social cost each time another decision must be made.

Multi-attribute utility measurement can spell out explicitly what the values of each participant (decision-maker, expert, pressure group, government, etc.) are, show how much they differ, and in the process can frequently reduce the extent of such differences. The exploitation of this technology permits regulatory or administrative agencies and other public decision-making organizations to shift their attention from specific actions to the values these actions serve and the decision-making mechanisms that implement these values. By explicitly negotiating about, agreeing on, and (if appropriate) publicizing a set of values, a decision-making organization can, in effect, inform those affected

by its decisions about its ground rules. This can often remove the uncertainty inherent in planning, and can often eliminate the need for costly, time-consuming, case-by-case adversary or negotiating proceedings. Thus, explicit social policies can be defined and implemented with more efficiency and less ambiguity. Moreover, such policies can easily be changed in response to new circumstances or changing value systems, and information about such changes can be easily, efficiently, and explicitly disseminated, greatly easing the task of implementing policy change.

The paper is structured around three examples. One is land use management; the specific example will be a study aimed at the decision problems of the California Coastal Commission. The decision-making body in this case is a regulatory agency exposed to a wide variety of social pressures from those with stakes in its actions.

The second example is concerned with administrative decision-making, specifically, with the process that the Office of Child Development of the U. S. Department of Health, Education, and Welfare used to develop its research program for the 1974 fiscal year.

The third example is more abstract; it concerns an attempt to develop a consensus among disagreeing experts on water quality, about a measure of the merits of various water sources for two purposes: the input, before treatment, to a public water supply, and an environment for fish and wildlife.

The focus of this paper is on planning. I do not understand the differences among evaluations of plans, evaluations of ongoing projects, and evaluations of completed projects; all seem to me to be instances of the same kind of intellectual activity. Multi-attribute utility measurement can and, I believe, should be applied to all three; the only difference is that in ongoing or completed projects there are more opportunities to replace judgmental estimates of locations on value dimensions with utility transforms on actual measurements--still subjective, but with firmer ground in evidence.

CONTENTS

FIGURES

TABLES

PREFACE

This report is a slightly edited version of a speech delivered at a conference of the International Institute of Applied Systems Analysis (IIASA) held in Laxenburg, Austria on 21 October 1975.

ACKNOWLEDGMENTS

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract No. N00014-76-C-0074.

I am grateful to Drs. Peter C. Gardiner, Marcia Guttentag, Michael F. O'Connor and Kurt Snapper for their permissions to review at length work for which they were wholly or partly responsible, and to Drs. Edith H. Grotberg and Ralph L. Keeney for their very helpful comments.

HOW TO USE MULTI-ATTRIBUTE UTILITY MEASUREMENT FOR SOCIAL DECISION MAKING

1.0 INTRODUCTION

Decisions do, and should, depend on values and probabilities--both subjective quantities. Public decisions, even more than other kinds, also should depend on values and probabilities. These quantities should be public, not only in the sense of being publishable, but also in the sense that the values, and perhaps the probabilities, that lie behind the decision should depend on some kind of social consensus, or at least on some kind of aggregation of individual views, rather than on any single individual's views.

The problem of obtaining such aggregate numbers differs for values and probabilities. A strong case can be made that probabilities should be generated out of data and expertise whenever both are available. Unless you happen to have a pocket calculator handy, your opinion about whether or not the natural logarithm of 222 is 540258 is not nearly so good as mine; I just calculated it. Considerations of social justice, every man's right to his own opinions, and the like, while never utterly irrelevant even to probabilities, become less and less important as differences in expertise become increasingly relevant. For that reason, this paper will ignore the many fascinating problems of combining or reconciling conflicting views about probabilities, and will deal only with the problem of public values.

As this paper later discusses in detail, the same point made in the preceding paragraph about probabilities applies

to values as well. Some aspects of value, specifically the location of the objects to be evaluated on the relevant dimensions of value, are also often matters of objective information, expertise, or some mixture of both. Yet most of us would agree that individuals are entitled to disagree about values and to have those disagreements respected and taken into account in public decision making. How can this be done?

Arrow's famous impossibility theorem (1951) has been interpreted by some as offering an answer: it can't. I cannot bring myself to take that answer very seriously, though I believe the theorem. Public decisions are made every day, and they do respond to individual differences in values in a crudely aggregative fashion. In my view, Arrow simply did not make sufficiently strong assumptions. For one thing, he worked with ordinal rather than cardinal utility; this paper takes cardinal utilities for granted. For another, he was unwilling to assume the interpersonal comparability of utilities. Yet, with or without axiomatic justification, we do in fact compare strengths of preference every day. That argument, carried to its extreme, would lead to the rather uninteresting idea of making social choices on the basis of averaged utilities of the people affected. We often do make social choices by mechanisms (e.g., voting) that have that flavor. But that is not the thrust of this paper.

The thrust of this paper is that a public value is a value assigned to an outcome by a public, usually by means of some public institution that does the evaluating. This amounts to treating "a public" as a sort of organism whose values can be elicited by some appropriate adaptation of the methods already in use to elicit individual values. From this point of view, the interest of the problem lies in finding the appropriate adaptation of those methods, an

adaptation that will take into account individual disagreements about values, individual differences in relevant expertise, existing social structures for making public decisions, and problems of feasibility.

The paper is structured around three examples. One is land use management; the specific example will be a study aimed at the decision problems of the California Coastal Commission. The decision-making body in this case is a regulatory agency exposed to a wide variety of social pressures from those with stakes in its actions. Because this public exposure to organized pressures is so explicit in this example, the paper will deal with it at great length; most of the issues that arise in this form of social decision-making arise also, often in subtler and more muted forms, in other decision contexts.

The second example is concerned with administrative decision-making; specifically, with the process that the Office of Child Development of the U. S. Department of Health, Education, and Welfare used to develop its research program for the 1974 fiscal year. It is the only one of the three examples in which the tools were used to make real decisions.

In a way, administrative decisions are misleading. The presence of a senior administrator with offical power to make the decisions suggests, incorrectly, that that administrator's values are being maximized by the decisions made. Seldom is the case that simple. For one thing, every boss has a boss, and attempts to take the values of his superiors into account in his own decisions. Moreover, every competent boss has a staff whose views he respects and whose values he regards as relevant, often more relevant than his own. Finally, administrative agencies often serve specific public constituencies, in addition to serving some abstract and

impersonal ideal of the public good. The fact that values differ from one staff member to another and from one constituency to another makes the case of the administrative decision-maker not greatly different from the case of the regulatory commission. By the time pressures from above and from below are taken into account, little room may be left for the administrator's own personal values.

The third example is more abstract; it concerns an attempt to develop a consensus, among disagreeing experts on water quality, about a measure of the merits of various water sources for two purposes: the input, before treatment, to a public water supply, and an environment for fish and wildlife. The experts were all involved in public decisions about water, but each worked in a different jurisdiction, so no need for consensus as a basis for decision existed. Still, agreed on measures of water quality for these purposes would be very useful.

The ideas presented in this paper are closely related to, and grow out of, those contained in Edwards (1971), Edwards and Guttentag (1975), and Edwards, Guttentag, and Snapper (1975). Conceptually, these discussions overlap. Also, they are closely related to those presented by Bauer and Wegener (1975), and indeed we, following their lead but not their footsteps, are also engaged in exploring the fusion of multi-attribute utility measurement with differential equation modeling as a tool for social planning. While this paper, being primarily concerned with existing applications, does not discuss that fusion, it may help the reader to keep its possibility in mind as a reason for this discussion of approaches to conflicting social values.

The focus of this paper is on planning. I do not understand the differences among evaluations of plans, evaluations of ongoing projects, and evaluations of completed

projects; all seem to me to be instances of the same kind of intellectual activity. Multi-attribute utility measurement can and, I believe, should be applied to all three; the only difference is that in ongoing or completed projects there are more opportunities to replace judgmental estimates of locations on value dimensions with utility transforms on actual measurements--still subjective, but with firmer ground in evidence.

The fundamental idea in a nutshell is this: Arguments over public policy typically turn out to hinge on disagreements about values. Such disagreements are often about degree, not kind; developed and developing nations may agree on the virtues both of increased industrialization and decreased degradation of the environment, but may differ about the relative importances of these goals. Normally, such disagreements are fought out in the context of specific decisions, over and over again, at enormous social cost each time another decision must be made. Multi-attribute utility measurement can spell out explicitly what the values of each participant (decision-maker, expert, pressure group, government, etc.) are, show how and how much they differ, and in the process can frequently reduce the extent of such differences. The exploitation of this technology permits regulatory or administrative agencies and other public decision-making organizations to shift their attention from specific actions to the values these actions serve and the decision-making mechanisms that implement these values. By explicitly negotiating about, agreeing on, and (if appropriate) publicizing a set of values, a decision-making organization can, in effect, inform those affected by its decisions about its ground rules. This can often remove the uncertainty inherent in planning, and can often eliminate the need for costly, time-consuming, case-by-case adversary or negotiating proceedings. Thus, explicit social policies can be defined and implemented with more efficiency and less ambiguity.

Moreover, such policies can easily be changed in response to new circumstances or changing value systems; and information about such changes can be easily, efficiently, and explicitly disseminated, greatly easing the task of implementing policy change.

2.0 A TECHNIQUE FOR MULTI-ATTRIBUTE UTILITY MEASUREMENT

Edwards (1971) has proposed the following technique for multi-attribute utility measurement based on extensive use of simple rating procedures. While it lacks the theoretical elegance of techniques proposed by, for example, Raiffa (1968, 1969) or Keeney (1972), it has the great advantage of being easily taught to and used by a busy decision-maker, or member of a decision-making staff organization. Moreover, it requires no judgments of preference or indifference among hypothetical entities. My experience with elicitation procedures suggests that such hypothetical judgments are unreliable and unrepresentative of real preferences; worse, they bore untutored decision-makers into either rejection of the whole process or acceptance of answers suggested by the sequence of questions rather than answers that reflect their real values, or both.

The basic idea of multi-attribute utility measurement is very familiar (see, for example, Raiffa, 1968). Every outcome of an action may have value on a number of different dimensions. The technique, in any of its numerous versions, is to discover those values, one dimension at a time, and then to aggregate them across dimensions using a suitable aggregation rule and weighting procedure. Probably the most widely used, and certainly the simplest, aggregation rule and weighting procedure consists of simply taking a weighted linear average; only that procedure will be discussed here. Theory, simulation computations, and experience all suggest that weighted linear averages yield extremely close approximations to very much more complicated non-linear and interactive "true" utility functions, while remaining far easier to elicit and understand. (See, for example, Wilks, 1938; Dawes and Corrigan, 1974; and Einhorn and Hogarth, 1975.)

2.1 Procedure

The technique consists of ten steps.

Step 1: Identify the person or organization whose utilities are to be maximized. If, as is often the case, several organizations have stakes and voices in the decision, they must all be identified. People who can speak for them must be identified and induced to cooperate.

Step 2: Identify the issue or issues (i.e., decision) to which the utilities needed are relevant. Depending on context and purpose, the same objects or acts may have many different values. In general, utility is a function of the evaluator, the entity being evaluated, and the purpose for which the evaluation is being made. The third argument of that function is sometimes neglected.

Step 3: Identify the entities to be evaluated. Formally, they are outcomes of possible actions. But in a sense, the distinction between an outcome and the opportunity for further actions is usually fictitious. The value of a dollar is the value of whatever one chooses to buy with it; the value of an education is the value of the things the educated person can do that he could not have done otherwise. Since it is always necessary to cut the decision tree somewhere, that is, to stop considering outcomes as opportunities for further decisions and instead simply to treat them as outcomes with intrinsic values, the choice of what to call an outcome becomes largely one of convenience. In practice, often it is sufficient to treat an action itself as an outcome. This amounts to treating the action as having an inevitable outcome, that is, of assuming that uncertainty about outcomes is not involved in the evaluation of that action. Paradoxically, this is frequently a good technique when the outcome is utterly uncertain, so uncertain that it is impractical or

not worthwhile to explore all its possible consequences in detail and assign probabilities to each.

When uncertainty is explicitly taken into account in social decision making, often the tool of choice for doing so is a set of scenarios, each with a probability. A scenario is simply a hypothetical future, organized around the stakes in the decision at hand and looking at the effect of various exogenous factors on their value. Considerable sophisticated experience in combining the use of scenarios with multi-attribute utilities exists, but is not yet available in print.

Step 4: Identify the relevant dimensions of value for evaluation of the entities. As Raiffa (1969) has noted, goals ordinarily come in hierarchies. But it is often practical and useful to ignore their hierarchical structure, and instead to specify a simple list of goals that seem important for the purpose at hand.

It is important not to be too expansive at this stage. The number of relevant dimensions of value should be modest, for reasons that will be apparent shortly. This can often be done by restating and combining goals, or by moving upward in a goal hierarchy. Even more important, it can be done by simply omitting the less important goals. There is no requirement that the list evolved in this step be complete, and much reason to hope that it will not be.

Step 5: Rank the dimensions in order of importance. This ranking job, like Step 4, can be performed either by an individual or by representatives of conflicting values acting separately or by those representatives acting as a group. I prefer to try group processes first, mostly to get the arguments on the table and to make it more likely that the participants start from a common information base, and

then to get separate judgments from each individual. The separate judgments will differ, of course, both here and in the following step.

Step 6: Rate dimensions in importance, preserving ratios. To do this, start by assigning the least important dimension an importance of 10. (We use 10 rather than 1 to permit subsequent judgments to be finely graded and nevertheless made in integers.) Now consider the next-least-important dimension. How much more important (if at all) is it than the least important? Assign it a number that reflects that ratio. Continue up the list, checking each set of implied ratios as each new judgment is made. Thus, if a dimension is assigned a weight of 20, while another is assigned a weight of 80, it means that the 20 dimension is 1/4 as important as the 80 dimension, and so on. By the time you get to the most important dimensions, there will be many checks to perform; typically, respondents will want to revise previous judgments to make them consistent with present ones. That's fine; they can do so. Once again, individual differences are likely to arise.

Step 7: Sum the importance weights, divide each by the sum, and multiply by 100. This is a purely computational step which converts importance wieghts into numbers that, mathematically, are rather like probabilities. The choice of a 1-to-100 scale is, of course, completely arbitrary.

At this step, the folly of including too many dimensions at Step 4 becomes glaringly apparent. If 100 points are to be distributed over a set of dimensions and some dimensions are very much more important than others, then the less important dimensions will have non-trivial weights only if there are not too many of them. As a rule of thumb, 8 dimensions is plenty, and 15 is too many. Knowing this, you will want at Step 4 to discourage respondents from being

too finely analytical; rather gross dimensions will be just right. Moreover, the list of dimensions may be revised later, and that revision, if it occurs, will typically consist of including more rather than fewer dimensions.

Step 8: Measure the location of each entity being evaluated on each dimension. The word "measure" is used rather loosely here. There are three classes of dimensions: purely subjective, partly subjective, and purely objective. The purely subjective dimensions are perhaps the easiest; you simply get an appropriate expert to estimate the position of the entity on that dimension on a 0-to-100 scale, where 0 is defined as the minimum plausible value and 100 is defined as the maximum plausible value. Note "minimum and maximum plausible" rather than "minimum and maximum possible." The minimum plausible value often is not total absence of the dimension.

A partly subjective dimension is one in which the units of measurement are objective, but the locations of the entities must be subjectively estimated.

A purely objective dimension is one that can be measured non-judgmentally, in objective units, before the decision. For partly or purely objective dimensions, it is necessary to have the estimators provide not only values for each entity to be evaluated, but also minimum and maximum plausible values, in the natural units of each dimension.

At this point we can identify a difference of opinion among users of multi-attribute utility measurement. Some (e.g. Edwards, 1971) are content to draw a straight line connecting maximum plausible with minimum plausible values and then to use this line as the source of transformed location measures. Others, such as Raiffa (1968), advocate the development of dimension-by-dimension utility curves.

Of various ways of obtaining such curves, the easiest way is simply to ask the respondent to draw graphs. The X-axis of each such graph represents the plausible range of performance values for the attribute under consideration. The Y-axis represents the ranges of values or desirabilities or utilities associated with the corresponding X values.

Strong reasons argue for the straight-line procedure whenever the underlying dimension is conditionally monotonic that is, either more is better than less or else less is better than more throughout the plausible range of the dimension regardless of locations on the other dimensions. These reasons essentially are that such straight lines will produce close approximations to the true value functions after aggregation over dimensions; correlations in excess of .99 are typical. Still, respondents are sometimes concerned about the non-linearity of their preferences, and may prefer to use the more complicated procedure. Additionally, preferences may not be monotone. Partly for these reasons, two of the three studies reported in this paper use non-linear value curves, though they avoid the elaborate techniques dependent on hypothetical indifference judgments that have often been proposed to obtain such curves.

A common objection to linear single-dimension value curves is that they ignore the economic law of diminishing returns. If you both prefer meat to drink and regard meat as more important than drink, and your utility function is linear with quantity of meat, you will keep on buying and perhaps consuming meat till you die of thirst. The objection is valid in some contexts, especially those in which the dimensions of value are separable, as they are in a commodity bundle, or those in which the set of available options is so rich that the dimensions might as well be separable. For contexts like those used as examples in this paper, the objection is irrelevant; linear single-dimension

value curves could have been used whenever conditional monotonicity applies in all three examples. The option of reducing less important dimensions to near-zero values did not exist.

In what sense, if any, are rescaled location measures comparable from one scale to another? The question cannot be considered separately from the question of what "importance," as it was judged at Step 6, means. Formally, judgments at Step 6 should be designed so that when the output of Step 7 is multiplied by the output of Step 8, equal numerical distances between these products correspond to equal changes in desirability. Careful instruction is usually needed to communicate this thought to respondents.

Step 9: Calculate utilities for entities. The equation is:

$$U_i = \sum_j w_j u_{ij},$$

remembering that $\sum_j w_j = 100$. U_i is the aggregate utility for the i^{th} entity. w_j is the normalized importance weight of j^{th} dimension of value, and u_{ij} is the rescaled position of the i^{th} entity on the j^{th} dimension. Thus, w_j is the output of Step 7 and u_{ij} is the output of Step 8. The equation, of course, is nothing more than the formula for a weighted average.

Step 10: Decide. If a single act is to be chosen, the rule is simple: maximize U_i. If a subset of i is to be chosen, then the subset for which $\sum_i U_i$ is maximum is best.

A special case arises when one of the dimensions, such as cost, is subject to an upper bound; that is, there is a budget constraint. In that case, Steps 4 through 10 should be done ignoring the constrained dimension. The ratios u_i/C_i, the cost of the i^{th} entity, should be chosen in

decreasing order of that ratio until the budget constraint is used up. (More complicated arithemtic is needed if programs are interdependent or if this rule does not come very close to exactly exhausting the budget constraint.) This is the only case in which the benefit-to-cost ratio is the appropriate figure on which to base a decision. In the absence of budget constraints, cost is just another dimension of value, entering into U_i with a minus sign, like other unattractive dimensions. In the general case, it is the benefit-minus-cost difference, not the benefit-over-cost ratio, that should usually control action.

An important caveat needs to be added concerning benefit-to-cost ratios. Such ratios assume that both benefits and costs are measured on a ratio scale, that is, a scale with a true zero point and ratio properties. The concepts both of zero benefit and of zero cost are somewhat slippery on close analysis. A not-too-bad solution to the problem is to assume that you know what zero cost means, and then attempt to find the zero point on the aggregate benefit scale. If that scale is reasonably densely populated with candidate programs, an approach to locating that zero point is to ask the decision maker, "Would you undertake this program if it had the same benefits it has now, but had zero cost?" If the answer is no, it is below the zero point.

The multi-attribute utility approach can easily be adapted to cases in which there are minimum or maximum acceptable values on a given dimension of value by simply excluding alternatives that lead to outcomes that transgress these limits.

2.2 Flexibilities of the Method

Practically every technical step in the preceding list has alternatives. For example, Keeney (1974) has proposed use of a multiplicative rather than an additive aggregation

rule. Certain applications have combined multiplication and addition. The methods suggested above for obtaining location measures and importance weights have alternatives; the most common is the direct assignment of importance weights on a 0-to-100 scale. (We consider this procedure inferior to the one described above, but doubt that it makes much practical difference in most cases.)

Because its emphasis is on simplicity and on rating rather than on more complicated elicitation methods, I call the above technique a Simple Multi-Attribute Rating Technique (SMART). I leave to critics the task of extending the acronym to show that its users are SMART-alecs.

2.3 Independent Properties

Either the additive or the multiplicative version of the aggregation rule assumes value independence. Roughly, value independence means that the extent of your preference for location a_2 over location a_1 of dimension A is unaffected by the position of the entity being evaluated on dimensions B, C, D, . . . Value independence is a strong assumption, not easily satisfied. Fortunately, in the presence of even modest amounts of measurement error, quite substantial amounts of deviation from value independence will make little difference to the ultimate number U_i, and even less to the rank ordering of the U_i values. [For recent discussions of the robustness of linear models, on which this assertion depends, see Dawes and Corrigan (1974) and Einhorn and Hogarth (1975).] A frequently satisfied condition that makes the assumption of value indeoendence very unlikely to cause trouble is conditional monotonicity; that is, the additive approximation will almost always work well if, for each dimension, either more is preferable to less or less is preferable to more throughout the range of the dimension that is involved in the evaluation for all available values

of the other dimensions. When the assumption of value independence is unacceptable even as an approximation, much more complicated models and elicitation procedures that take value dependence into account are available.

A trickier issue than value independence is what might be called environmental independence. The traffic congestion caused by a coastal development is extremely likely to be positively correlated with the number of people served by the development. Yet these two dimensions may be value-independent; the correlation simply means that programs with both little traffic congestion and many people served are unlikely to present themselves for evaluation.

Violations of environmental independence can lead to double counting. If two value dimensions are perfectly environmentally correlated, only one need be included in the evaluation process. If both are included, care must be taken to ensure that the aggregate importance weight given to both together properly captures their joint importance. For example, if number of people served and traffic congestion were perfectly environmentally correlated and measured on the same scale after rescaling, if they had equal weights, and if one entered with a positive sign and the other with a negative sign into the aggregation, the implication would be that they exactly neutralized each other, so that any feasible combination of these two variables would be equivalent in value to any other feasible combination. The decision maker is unlikely to feel that way, but may have trouble adjusting his importance weights to reflect his true feelings. His life could be simplified by redefining the two dimensions into one, e.g., number of people served, taking into consideration all that that entails with respect to traffic.

The problem is trickier if the environmental correlation is high but not perfect. But the solution remains the same:

Try, whenever possible, to define or redefine value dimensions in order to keep environmental correlations among them low. When that cannot be done, check the implications of importance weights and location measures assigned to environmentally correlated dimensions to make sure that their aggregate weight properly reflects their aggregate importance.

Similar comments apply, though transparent examples are harder to construct, when the sign of the environmental correlation and the signs with which the dimensions enter into the aggregate utility function are such that double counting would over- rather than under-emphasize the importance of the aggregate of the two dimensions.

A final technical point should be made about environmental correlations.[1] In general, if you must choose one entity from all the possibilities, the correlation between the dimensions will be large and negative. In the technical language of decision theory, the point is simply that the undominated set of entities (i.e. the contending entities) must lie on the convex boundary and so are necessarily negatively correlated with one another. This point becomes much less significant when one is selecting a number of entities rather than just one, since the selection of each entity removes it from the choice set, redraws the convex boundary of remaining entities, and probably thus reduces the negative correlation.

Unfortunately, the higher the negative environmental correlation among value dimensions, the less satisfactory becomes the use of the value independence assumption as an

---

[1] I am grateful to David Seaver, who first called the issue discussed in the following paragraphs to my attention.

approximation when value correlations are actually present. At present, I know of no detailed mathematical or simulation study of the effect of size of the environmental correlation on acceptability of the value-independence approximation. This question is likely to receive detailed examination in the next few years.

3.0 ILLUSTRATIVE APPLICATIONS OF THE TECHNIQUE

3.1 Example 1: Land Use Regulation by the California Coastal Commission[2]

Prior to 1972, two hundred separate entities--city, county, state, and federal governments, agencies and commissions--regulated the California coast. The citizens of California, in reviewing the performances of these two hundred entities, were apparently dissatisfied, and in a voter-sponsored initiative during the general election of 1972, the voters approved legislation placing coastal zone planning and management under one state commission and six regional commissions. In passing the Coastal Zone Conservation Act by 55% of the vote, the voters established decision makers with ultimate authority (other than appeal to the courts) to preserve, protect, restore, and enhance the environment and ecology of the state's coastal zone.[3]

The coastal zone is defined in the Act as the area between the seaward limits of state jurisdiction and 1,000 yards landward from the mean high-tide line. Any plan for development within the coastal zone must be approved by the appropriate regional commission before it can be carried out. Disapprovals can be appealed to the state commission and then to the courts if necessary. (Development permits are similar to other types of building permits and authorize only the specific activities named.)

The South Coast Regional Commission (Region V) comprising Los Angeles and Orange counties is one of the six

---

[2] This example, based on Dr. Peter Gardiner's Ph.D. thesis (Gardiner, 1974), has also been discussed at length in Gardiner and Edwards (1975).

[3] California Coastal Zone Conservation Act, 1972.

regional commissions. Los Angeles county is heavily urbanized and in 1970 contained 35% of the total state population and 41% of the state's coastal county population. Los Angeles county includes the coastal cities of Long Beach, Redondo Beach, Hermosa Beach, Manhattan Beach, Los Angeles (Venice and the harbor area), Santa Monica, and unincorporated county areas such as Marina del Rey. These cities and areas all contain portions of the coastal zone that are under the control of the Region V Commission. Approximately one billion dollars worth of development was authorized in the first year of the commission's activities and over 1,800 permits were acted upon. A backlog of as many as 600 permit requests awaiting action has existed. The evaluation and decision-making tasks that confront the Region V Commission members are important, far-reaching, difficult and controversial.

Although the Act specified that certain attributes should be considered in making evaluations, it fails to specify just how they are supposed to enter into the evaluation process. Nor does the Act specify how the Commissioners are to balance the conflicting interests affected by their decisions. In effect, the Act implies that individual commissioners assigned to the Commission will represent the interests of all affected parties with respect to the coastal zone in Region V. How this is to be accomplished is left unspecified. In practice, attempts to include the preferences and value judgments of interested groups and individuals occur when the Commission holds public advocacy hearings on permit requests. Under these procedures, opposing interest groups express their values and viewpoints as conclusions, often based on inconsistent sets of asserted facts or no facts at all, in the form of verbal and written presentations at the open hearings.

3.1.1 <u>Procedure</u> - Fourteen individuals involved in coastal zone planning and decision making agreed to participate in this study. Included were two of the current Coastal Commissioners for Region V, a number of active conservationists, and one major coastal zone developer. The purpose of this study was to test the consequences of using multi-attribute utility measurement processes by having participants in or people close to the regulatory process with differing views make both individual and group evaluations of various proposals for development in a section of the California coastal zone. Evaluations were made both intuitively and by constructing multi-attribute utility measurement models.

To provide a common basis for making evaluations, a sample of fifteen hypothetical but realistic permit requests for development were invented. The types of permits were limited to those for development of single-family dwellings, duplex, triplex, or multi-family dwellings (owned or for renting). Dwelling unit development (leading to increased population density) is a major area of debate in current coastal zone decision making. Most permit applications submitted to the Region V Commission thus far fall into this class. Moreover, permits granted in this class will probably generate further permit requests. Housing development tends to bring about the need for other development in the coastal zone such as in public works, recreation, transportation, and so on. The permit applications provided eight items of information about the proposed development that formed the information base on which subjects were asked to make their evaluations. These eight items were abstraced from actual staff reports currently submitted to the Region V coastal commissioners as a basis for their evaluations and decision making on current permit applications. The Commissioners' staff reports do have some additional information such as the name of the applicant and so on, but the following items are crucial for evaluation:

1. Size of development. The number of square feet of the coastal zone taken up by the development.

2. Distance from the mean high-tide line. The location of the nearest edge of the development from the mean high-tide line measured in feet.

3. Density of the proposed development. The number of dwelling units per acre for the development.

4. On-site parking facilities. The percentage of cars brought in by the development that are provided parking space as part of the development on-site.

5. Building height. The height of the development in feet (17.5 feet per story).

6. Unit rental. The dollar rental per month (on the average) for the development. If the development is owner-occupied and no rent is paid, an equivalent to rent is computed by taking the normal monthly mortgage payment.

7. Conformity with land use in the vicinity. The density, measured on a five-point scale from much less dense to much more dense, of the development relative to the average density of adjacent residential lots.

8. Esthetics of the development. A rating on a scale from poor to excellent.

Each of the invented permits was constructed to report a level of performance for each item. They were as realistic

as possible and represented a wide variety of possible permits.

Each subject answered seven questionnaires. In general, the participants had 5 days to work on each of the questionnaires. In the process of responding to the seven questionnaires each subject (1) categorized himself/herself on an eleven-point continuum that ranged from very conservationist-oriented to very development-oriented; (2) evaluated intuitively (holistically) 15 sample development permit requests by rating their overall merit on a 0-to-100 point worth scale; (3) followed the steps of multi-attribute utility measurement outlined previously and in so doing constructed individual and group value models;[4] and (4) reevaluated the same 15 sample permit requests intuitively a second time. Subjects did not know that the second batch of permits was a repetition of the first.

The location of the proposed developments was Venice, California, which is geographically part of the city of Los Angeles, located between Santa Monica and Marina del Rey. Venice has a diverse population and has been called a microcosm, a little world epitomizing a larger one (Torgerson, 1973). In many ways, Venice presents in one small area instances of all the most controversial issues associated with coastal zone decision making.

---

[4]The evaluation and decision making in this study are assumed to be riskless. Decisions involving permit requests, by the nature of the permits themselves, suggest that the consequences of approval or disapproval are known with certainty. The developer states on his permit what he intends to do if the permit is approved and is thereby constrained if approval is granted. If the request is disapproved, there will be no development, unless the present or subsequent owner of the land presents a new or revised request. Revision of both permit requests to meet Commission objectives often occurs, both before and after the original hearing. In that sense, the Commission's decisions are risky, but the possibility was omitted from the present study.

After the initial questionnaire, in which the subjects categorized themselves according to their views about coastal zone development, the fourteen individuals were divided into two groups. Group 1 was the eight more conservationist-minded subjects and Group 2 was the other six subjects whose views, by self-report, ranged from moderate to strongly pro-development.

In both the intuitive evaluation and multi-attribute utility measurement tasks, the subjects reported no major difficulty in completing the questionnaires. An example of one participant's value curves and importance weights is shown in Figure 1. The abscissae represent the natural dimension ranges and the ordinates represent value ranging from zero to one hundred points. Although the value curves shown are all monotone and could therefore be linearly approximated as indicated earlier, eleven of the fourteen subjects produced at least one non-monotone value curve. Accordingly, this study used the actual value curves for each subject rather than the linear approximation.

To develop group intuitive ratings and group value models, each individual in a group was given, through feedback, the opportunity of seeing his group's initial responses on a given task (intuitive ratings, importance weights, etc.) and of revising his own judgments. These data were fed back in the form of group means. Averaging individual responses to form group responses produced the results shown in Table 1. Table 1 shows in column 2 test-retest holistic evaluations of the 15 sample permits. These correlations are computed by taking the mean group ratings for each permit on the initial (test) intuitive evaluation and the second (retest) intuitive evaluation. The test holistic-SMART evaluation correlations are computed by comparing a group value model's ratings of the 15 sample permits with the group's initial intuitive evaluations. The

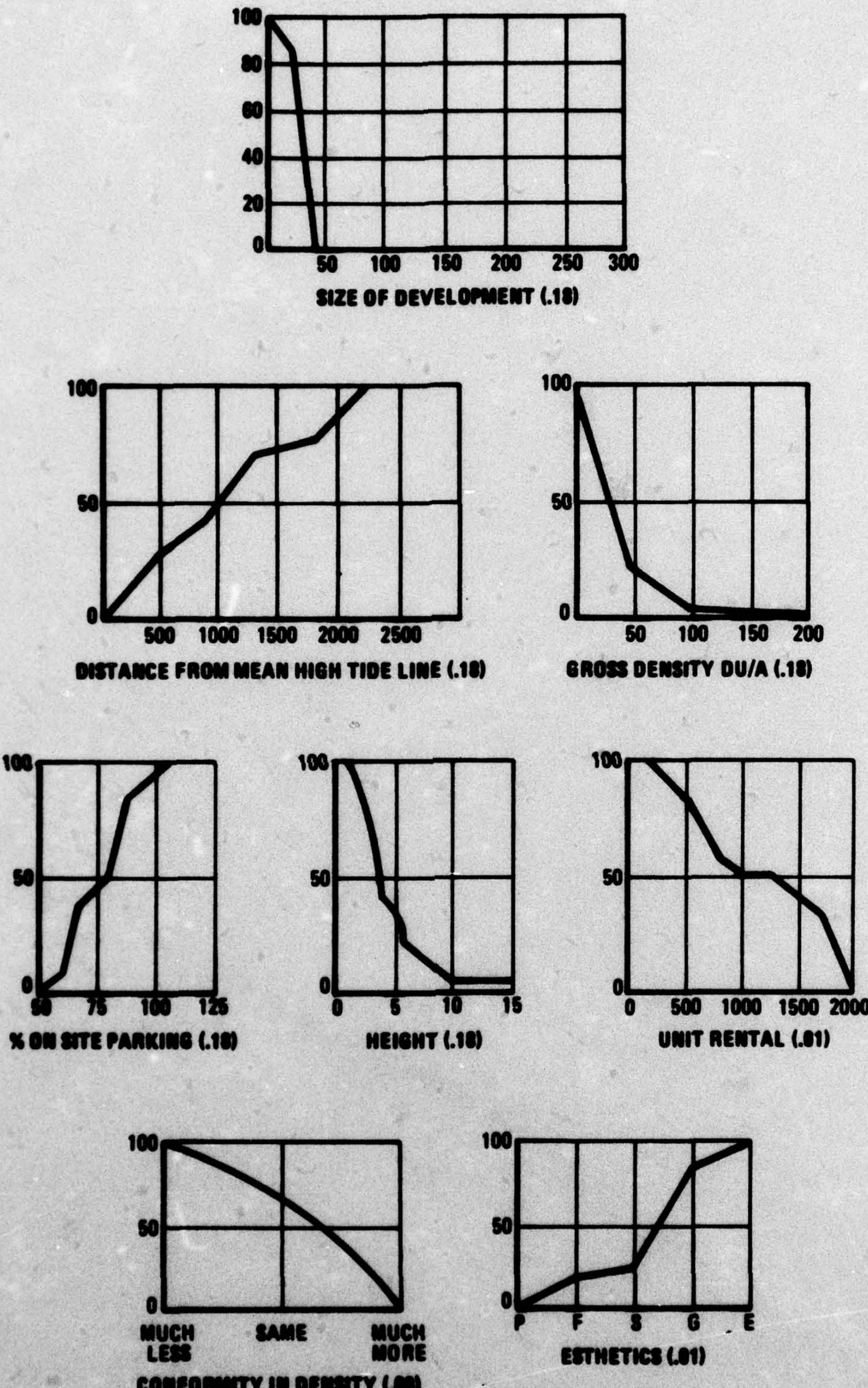


Figure 1

AN EXAMPLE OF VALUE CURVES AND IMPORTANCE WEIGHTS (IN PARENTHESES) FOR PERMIT REQUEST DIMENSIONS.

GROUP	EVALUATIONS (RELIABILITY)	TEST HOLISTIC-SMART EVALUATIONS	RETEST HOLISTIC-SMART EVALUATIONS
1	0.949	0.944	0.917
2	0.867	0.665	0.873

TABLE 1: GROUP PRODUCT MOMENT CORRELATIONS

group value model is found by computing the mean importance weights and mean value curves for the group and then evaluating each permit using the group's value model. The retest holistic-SMART evaluation correlations are similar except that the second intuitive evaluation is used.

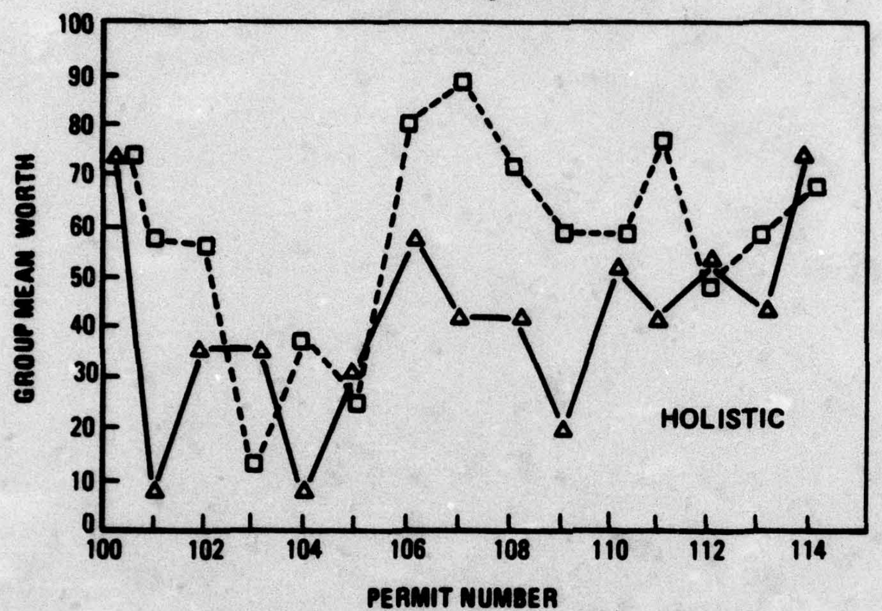
As can be seen from Table 1. each group's value model, constructed according to the procedures of multi-attribute utility measurement, has apparently "captured" the holistic evaluations of the group reasonably well. The interesting question is then, "What is the effect of using a group's value models vs. a group's intuitive evaluation?"

To answer this question, a two-way analysis of variance of permit worths was conducted. The independent

variables were groups and permit requests. These results indicate that the two groups initially (i.e., by holistic intuitive evaluations) represented differing viewpoints (i.e., were drawn from differing populations) although the differences were not dramatic. Substantial percentages of variance were accounted for both by group main effects and by permit-group interactions for the first-test holistic evaluations. Results for the retest were similar. Both findings indicate differing viewpoints between the two groups. The main effect could be caused, however, by a constant evaluation bias alone. The key indication of differing viewpoints is the interaction term. The use of each group's value model evaluations instead of their intuitive evaluations causes the percent of variance accounted for by the interaction to drop from 12% to 2%. Figure 2 shows this difference dramatically. The multi-attribute utility technique has turned modest disagreement into substantial agreement.

Why? Here is a plausible answer. When making holistic evaluations, those with strong points of view tend to concentrate on those aspects of the entities being evaluated that most strongly engage their biases. The multi-attribute procedure does not permit this; it separates judgment of the importance of a dimension from judgment of where a particular entity falls on that dimension. These applications varied on eight dimensions relevant to the environmentalists-versus-builders arguments. While these two views may cause different thoughts about how good a particular level of performance on some dimensions may be, evaluation on other dimensions will be more or less independent of viewpoint. Agreement about those other dimensions tends to reduce the impact of disagreement on controversial dimensions. That is, multi-attribute utility measurement procedures do not foster an opportunity for any one or two dimensions to become so

PANEL A: Mean worth for each group as a function of permit number based on holistic judgments of worth.



PANEL B: Mean worth for each group as a function of permit number based on each judge's SMART judgments.

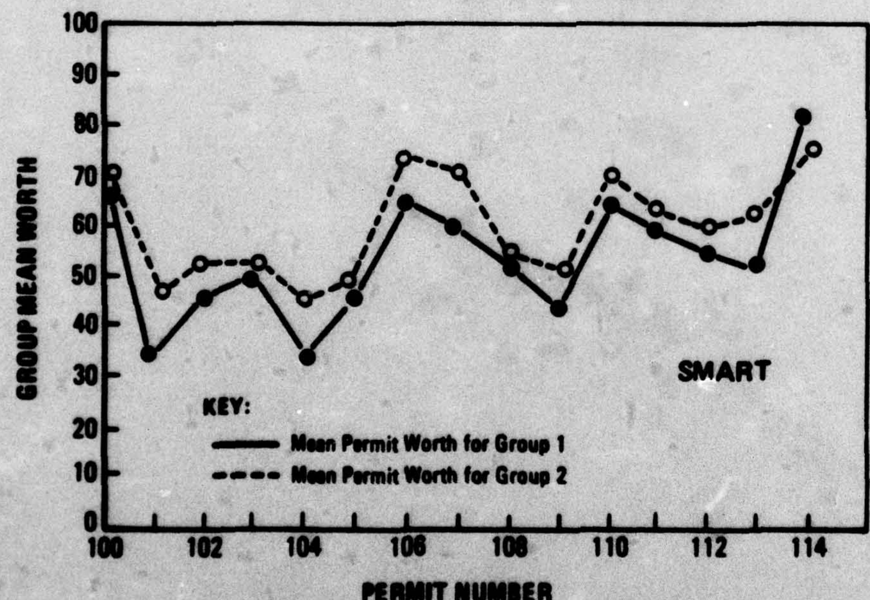


Figure 2
SMART-FOSTERED AGREEMENT.

salient that they emphasize existing sources of conflict and disagreement. Multi-attribute utility measurement cannot and should not eliminate all disagreement, however; such conflicts are genuine, and any value measurement procedure should respect and so reflect them. Still, in spite of disagreement, social decisions must be made. How?

I distinguish between two kinds of disagreements. Disagreements at Step 8 seem to me to be essentially like disagreements among different thermometers measuring the same temperature. If they are not too large, one has little compunction about taking an average. If they are, then one is likely to suspect that some of the thermometers are not working properly and to discard their readings. In general, I think that judgmentally determined location measures should reflect expertise and, typically, I would expect different value dimensions to require different kinds of expertise and therefore different experts. In some practical contexts, one can avoid the problem of disagreement at Step 8 entirely by the simple expedient of asking only the best available expert for each dimension to make judgments about that dimension.

Disagreement at Steps 5 and 6 are another matter. These seem to me to be the essence of conflicting values, and I wish to respect them as much as possible. For that reason, the judges who perform Steps 5 and 6 should be either the decison-maker(s) or well-chosen representatives. Considerable discussion, persuasion, and information exchange should be used in an attempt to reduce the disagreements as much as possible. At the least, this process offers a clear definition of the rules of debate and an orderly way to proceed from information and data, to values, to decisions.

Even this will seldom reduce disagreements entirely, however. The next two examples will suggest ways to proceed further.

3.1.2 Comment: A public technology for land use management - I conclude this example with a rather visionary discussion of how agencies responsible for land use management could carry out the task of land use management by fully exploiting SMART or some similar value measurement technique.

The statutes would define, at least to some degree, the appropriate dimensions of value, as they do now. They might, but probably should not, specify limits on the importance weights attached to these dimensions. They might and perhaps should specify boundaries beyond which no value could go in the undesirable direction.

The main functions of the regulatory agency would be four: 1) to specify measurement methods for each value dimension (with utility functions or other methods for making the necessary transformations at Step 8); 2) to specify importance weights; 3) to define one or more bounds not specified by statute on specific dimensions; and 4) to hear appeals.

The regulatory agency could afford to spend enormous amounts of time and effort on its first two functions, specification of measurement methods and of importance weights. Value considerations, political considerations, views of competing constituencies and advocates, the arts of logrolling and compromise--all would come into play. Public hearings would be held, with elaborate and extensive debate and full airing of all relevant issues and points of view.

The regulatory agency would have further responsibilities in dealing with measurement methods for wholly or partly subjective value dimensions. Since such measurements must be judgments, the regulatory agency must make sure that

the judgments are impartial and fair. This could be done by having staff members make them, or by offering the planner a list of agency-approved impartial experts, or by mediating among or selecting from the conflicting views of experts selected by those with stakes in the decision, or by some combination of these methods. I consider the first two of these approaches to be most desirable, but recognize that the third or fourth may be inevitable.

The reason why the costs of prolonged and intensive study of measurement methods and of importance weights could be borne is that they would recur infrequently. Once agreed-on measurement methods and importance weights had been "hammered out," most case-by-case decisions would be automatically made by means of them. Only in response to changed political and social circumstances or changed technology would reconsideration of the agreed-on measurement methods and importance weights be necessary, and even such reconsiderations would be likely to be partial rather than complete. They would, of course, occur; times do change, public tastes and values change, and technologies change. Those seeking appropriate elective offices could campaign for such changes; an election platform consisting in part of a list of numerical importance weights would be a refreshing novelty!

The decision rules would, of course, be public knowledge. That fact probably would be the most cost-saving aspect of this whole approach. Would-be developers and builders would not waste their time and money preparing plans that they could easily calculate to be unacceptable. Instead, they would prepare acceptable plans from the outset. Once a plan had been prepared and submitted to the regulatory agency, its evaluation would consist of little more than a check that the planner's measurements and

arithmetic had been done correctly. Delay from submission to approval need be no more than a few days.

Changes in the decision rules can be and should be as explicit as the rules themselves. Such explicitness would permit regulators and those regulated alike to know exactly what current regulatory policies are and, if they have changed, how and how much. Such knowledge would greatly facilitate both enlightened citizen participation in deciding on policy changes and swift, precise adaptation of those regulated to such changes once they have taken effect.

In short, multi-attribute utility measurement allows value conflicts bearing on social decisions to be fought out and resolved at the level of decision rules rather than at the level of individual decisions. Such decision rules, once specified, define and thus remove nearly all ambiguity from regulatory policy without impairing society's freedom to modify policies in response to changing conditions. Possible savings in financial and social costs, delays, frustrations, and so on are incalculable, but cost reduction in dollars alone could be 90% or more.

The idea of resolving value conflicts at the level of decision rules rather than at the level of individual decisions may have the potential of revolutionary impact on land use management and many other public decision contexts as well. Any new idea is bound to be full of unexpected consequences, traps, and surprises. For a while, therefore, the wise innovator would want to run old and new systems in parallel, compare performance of the two, and build up experience with the new system. A good mechanism might be to define an upper and lower bound, with automatic acceptance above the upper bound, automatic rejection below the lower one, and hearings in between. That would provide a

convenient administrative device for operation of such parallel procedures. Initially the upper bound could be very high and the lower bound very low so that most cases would fall in between and be handled by the traditional hearing mechanism. A candidate number for the lower bound, at least initially, is the utility of the do nothing (i.e., status quo) alternative, for obvious reasons. If what the applicant wants is not clearly better than the status quo, why does he deserve a hearing? As experience and confidence in the multi-attribute utility measurement system develop, the two bounds can be moved toward each other, so that more and more cases are handled automatically rather than by means of hearings. This process need work no hardship on any rejected applicant; he can always appeal, accepting the delays, costs, and risk of losing implicit in the hearing process rather than the cost of upgrading his plan. And the regulatory agency, by moving the boundaries, can in effect control its case load and thus gradually shorten the frequently inordinate delays of current procedures.

At present, I know of no public context in which even limited experimentation with these methods is occurring. But I have hopes.

3.2 Example 2: Planning a Government Research Program

The Office of Child Development (OCD) of the U.S. Department of Health, Education, and Welfare has a variety of responsibilities. Perhaps the largest is the operation of Project Head Start, a very large program for facilitating the development of pre-school children that is not included in this example. But it also sponsors a research program concerned with methods for promoting child welfare, for dealing with specific problems of children, and the like.

In the fall of 1972, OCD was faced with the task of planning its research program for fiscal 1974, which began on July 1, 1973. Guidance from the Department of Health, Education, and Welfare indicated that this research program, unlike its predecessors, would have to be justified by means of some assessment of its costs and benefits. While OCD staff members knew how to assess the cost of a research program, they had considerable difficulty in thinking about how to assess its benefits in quantitative form. So a team consisting of Marcia Guttentag, Kurt Snapper, and me were brought in as consultants, to work primarily with John Busa of OCD on the analysis. Dr. Guttentag is an expert at social psychological work in general and evaluation research in particular. Dr. Snapper moved to Washington at the beginning of 1973 to work on the OCD project full-time. Without his energy, imagination, and adaptability, the project could never have reached its successful conclusion. A fuller report of this project has been published by Guttentag and Snapper (1974).

3.2.1 <u>Procedure</u> - The ten-step process specified earlier in this paper was used. Initially, we assumed that the organization whose utilities were to be maximized was OCD. We later learned that this was a considerable oversimplification. Initially, we assumed that the entities to be evaluated were proposed research programs; this initial assumption, too, turned out to be excessively simplistic.

<u>Step 4</u>. To carry out Step 4, OCD assembled for two days a face-to-face group of some 15 people, consisting of OCD administrators and staff, both from Washington and from OCD field offices all over the country, plus several academic experts on child development. At my insistence, the value dimensions were segregated into two lists, one concerned with benefits to children and families and the

other concerned with benefits to OCD as an organization. My reason for the distinction is that in previous applications of the method, I had found that dimensions that were in fact concerned with organizational survival and growth were frequently encoded in language that sounded as though they referred to fulfillment of the organizational mission; organizations are often unwilling to admit the importance of survival and growth in controlling their decisions. Thus, for example, a dimension that in fact was, "Enhance the impact of OCD on federal programs related to child health" might appear as, "Promote child health." It seemed to me that a clearer picture of OCD's actual values could be obtained if the values associated with organizational survival and growth were segregated from those concerned with fulfillment of its mission, so that each class of values could be dealt with separately.

Initial lists of value dimensions (called goals or criteria to facilitate communication with the respondents) in each of the two groups were elicited by inviting the participants to state those goals; each list ended up with about 35-40 goals on it. A major task was then to pare the lists. Early eliminations were easy because some of the goals were simply restatements of others in slightly different language or because everyone agreed that a particular goal was not important enough to be worth considering or was not relevant to designing a research program. Later, more difficult paring of the lists was accomplished by having each participant rank-order the importances of the goals in each list separately, and then proposing goals that were low on most rank orders for deletion. This process produced many deletions; more important, it produced extremely searching and sophisticated discussions of just what each goal meant, how it related to other goals, and what sort of research or other action might serve it. These discussions combined with

the social effects of face-to-face interaction to produce considerably more agreement about the meanings of the various goals and their relative importances than would have occurred otherwise, though, of course, the agreement was very far from complete.

Steps 5 and 6. Each participant in the process was then asked to perform Steps 5 and 6 individually. All 13 forms were returned with usable ratings. A few more goals were eliminated on the basis of these ratings, essentially on the argument that they contributed 5% or less to total importance, and respondents seemed rather well-agreed on their low level of importance. Of course, with all the low-rating dimensions eliminated at various stages along the way, the remaining high-importance dimensions showed considerable interpersonal disagreement. Careful analysis showed that disagreement was not systematically related to the race, sex, or organizational locus of the respondent.

The Acting Director of OCD assigned final importance weights, mostly in good agreement with the means of the 13 respondents. He also made judgments relating importance weights across the two lists, values to children and families, and values to OCD. These judgments permitted the consolidation of those two lists, with their separate importance weights, into one list:

Criterion A (Importance weight = .007)

The extent to which a recommended activity is likely to foster service continuity/coordination and elimination of fragmentation, or is likely to contribute to this goal.

Criterion B (Importance weight = .145)

The extent to which a recommended activity represents an investment in a prototypical and/or high-leverage activity, or is likely to contribute to the development of prototypical/high-leverage programs.

Criterion C (Importance weight = .061)

The extent to which a recommended activity increases or is likely to contribute to an increase in families' sense of efficacy and their ability to obtain and use resources necessary for the healthy development of children.

Criterion D (Importance weight = .052)

The extent to which a recommended activity is likely to increase the probability that children will acquire the skills necessary for successful performance of adult roles, or is likely to contribute to that goal.

Criterion E (Importance weight = .036)

The extent to which a recommended activity is likely to contribute to making the public and institutions more sensitive to the developmental needs of children.

Criterion F (Importance weight = .048)

The extent to which a recommended activity is likely to promote the individualization of services or programs, or is likely to contribute to this goal.

Criterion G (Importance weight = .043)

The extent to which a recommended activity is likely to stimulate the development of pluralistic child care delivery systems that provide for parental choice, or is likely to contribute to the expansion of such systems.

Criterion H (Importance weight = .014)

The extent to which a recommended activity is likely to promote self-respect and mutual regard among children from diverse racial, cultural, class, and ethnic backgrounds, or is likely to contribute to this goal.

Criterion I (Importance weight = .009)

The extent to which a recommended activity is likely to result in effective interagency coordination at federal, state, and local levels, or is likely to contribute to this goal.

Criterion J (Importance weight = .160)

The extent to which a recommended activity is consonant with administration and departmental policies and philosophy, or reflects prevailing public and social thinking.

Criterion K (Importance weight = .120)

The extent to which a recommended activity is likely to make public leadership more sensitive to the needs of children.

Criterion L (Importance weight = .145)

The extent to which a recommended activity is likely to influence national child care policy in a positive way.

Criterion M (Importance weight = .032)

The extent to which a recommended activity is capable of rational explication, that is, the extent to which it represents a logical extension of past results and conclusions, is indicated on theoretical grounds, or fulfills prior commitments.

Criterion N (Importance weight = .129)

The extent to which a recommended activity is likely to produce tangible, short-term results, that is, the extent to which it is likely to produce or contribute to the production of solid conclusions, benefits, or results within a relatively short period of time.

The dimensions had acquired considerably more careful definitions along the way. Of the five criteria receiving weights of .10 or more, four came from the values to OCD rather than the values to children and families list. And even Criterion B, which, in fact, was on the values-to-children-and-families list, might have been on the other list as well. These findings should be no surprise to students of administrative and bureaucratic decision-making. They should, however, give researchers reason to pause for thought. Especially interesting was the fate of one goal that had appeared on the first list of values to children and families: "Contribute to knowledge expansion and/or use of knowledge for program

planning." This was easily eliminated as relatively unimportant. At the time, I found its elimination baffling, since I had been told that the goal of the exercise was to evaluate research proposals. As it turned out, this was not the goal of the exercise; I had failed to perform Step 3 properly. Moveover, OCD is an organization interested in applying knowledge to problems. Its programs are mostly action-oriented. New knowledge is important only if it can lead to more effective action. Consequently, the value of new knowledge should derive from its contribution to action goals. Thus, the elimination of a goal that in effect valued knowledge for its own sake was consistent with the basic mission and value structure of OCD.

Step 3. When this project started, I had supposed that OCD received a flow of research proposals, and that we were to develop a method of deciding which ones to implement or fund. That was naive of me! Actually, OCD projects start as statements of research priorities or as Requests for Proposals. The question of what we were trying to evaluate might have been much better handled if I had understood better at the time how the process by which OCD generates its research program differs from the process by which some other HEW agencies, such as the National Institutes of Health, generate theirs.

Still, we supposed that we were trying to evaluate specific research activities. So we set out to create a list of activities to evaluate. Suggested research projects came from many sources. Major reports to OCD and HEW were summarized, and their recommendations were restated, where appropriate, as research projects. Recommendations were obtained from many members of the OCD Staff, from the Office of Assistant Secretary for Planning and Evaluation in HEW,

and many other interested government and private groups. Several hundred recommendations were assembled, combined, and refined as a result of this process. For specificity, a proposed duration and cost was attached to each. Most of these were in the range of 1 to 3 years and $50,000 to $1,500,000.

Step 8. Informal screening was used to reduce the output of Step 3 to a smaller and more manageable set; ultimately, 56 research recommendations were carried through the entire analysis. Each of these 56 recommendations was independently scaled on each of the 13 dimensions of value by three members of the OCD staff, 56 x 13 x 3 = 2184 judgments in all. Inter-judge reliability was generally quite good, considerably higher than it had been for the importance weights, and quite high enough so that we had no compunction about taking the average over the 3 judges as the scale value for each research project on each dimension. The projects scattered out well over each dimension. For example, for dimension H the range was from 880 to 260; for dimension G the range was from 470 to 25.

Step 9. Calculation of utilities for each recommendation required no more than multiplication and addition. The range of aggregate utilities for the 56 research recommendations was from about 550 to about 200, and the distribution was well spread out over that range; the mean was 369 and the standard deviation of the the 56 utility values (on the scale) was about 71.5. For convenience, the scale was stretched out by a linear transformation so that the lowest aggregate utility was 0 and the highest was 1000. On this new scale, the mean was 483 and the standard deviation was 204.

The next step, since we wanted to look at benefit-to-cost ratios, was to see if the utility scale had a locatable true zero point. The Acting Director of OCD was asked whether there were any projects on the list that he would not wish to have OCD sponsor even if they were free. There were 10 such projects. A cutting score of 295 (on the 0-to-1000 rescaled utility function) identified them with only one inversion. So 295 was adopted as the zero point of the 0-to-1000 utility scale (which thus became a 0-to-705 scale); projects falling below that score were dropped from consideration, and benefit-to-cost ratios were calculated for the rest. Ordering in cost-benefit ratio of course differed from ordering in benefits alone.

Step 10. Our failure to perform Step 3 properly now caught up with us. The process by which we had produced proposed research topics was casual and ad hoc, and the results showed it. The proposed topics did not cover all important substantive areas of research on child development and were not well formulated with respect to the topics they did cover. Moreover, by this time we had a somewhat better understanding of what role the evaluative machinery we had developed could serve. It was not well designed to evaluate specific research projects, but it could evaluate higher-order questions having to do with directions in which research programs might go.

Step 3 again. Working with OCD scientists, we jointly developed a comprehensive taxonomy of research areas, taking into account those that had been omitted as well as those that had been included in the previous list of research projects. This produced a short list of general research foci that subsumed most of the previously generated specific projects.

Step 10 again. Using only the five value dimensions with highest weights, each general research focus was evaluated. A rough rule-of-thumb was proposed: each research focus should receive a portion of the available funds proportional to its utility. The acting Director of OCD, with value dimensions in hand but initially without utilities of the research areas, made a tentative allocation of funds. This allocation was compared with the result of the rule-of-thumb. The relationship was close, though not perfect. So the Acting Director reduced the funding of areas that received too much by that rule-of-thumb and increased the funding of areas that received too little. A comparison of the 1973 with the 1974 research budget allocations clearly shows that changes did occur in these directions and in amounts close to those suggested by the rule-of-thumb.

3.2.2 Conclusion - In retrospect, the most serious deficiency of the procedure was failure to perform steps 1,2, and 3 in time. Step 1 caused difficulties; not only the values of OCD, but also those of reviewing organizations within DHEW were relevant and should have been ascertained. But the most important failure was that the procedures for performing Step 3 were hasty and ad hoc, and resulted in unsatisfactory lists of research recommendations. This failure ultimately forced the decision process to a much higher level of abstraction, at which broad research areas rather than specific projects were evaluated.

While this was not what we had originally had in mind, it may have served OCD well. The value dimensions originally elicited from OCD staff members and others were not particularly appropriate to evaluating specific research projects. They did not address such questions as the feasibility of the project, the extent to which it related to what had

already been done, the extent to which it advanced knowledge in some significant areas, and so on. On the other hand, those dimensions did address the question, "What do OCD staff members value?" So they are more appropriate for broad programmatic guidance than for evaluating specific projects. It would have been an interesting and valuable exercise in using hierarchical value structures to develop a second evaluative mechanism suitable for evaluating specific responses to statements of OCD research priorities or Requests for Proposals. Such an evaluative mechanism should measure congruence of the responses with OCD's broad values as reflected in the requests that stimulated them while at the same time measuring the congruence of those responses with the general criteria one uses to evaluate social-science research projects. But we were not asked to do that.

The method used to obtain value dimensions and importance weights seemed to work well in a technical sense. The extensive use of group discussion, interspersed with ratings and re-ratings, considerably enhanced OCD's awareness both of its own value and of value conflicts within its staff, and in the process did much to reduce those conflicts. In retrospect, this was the most important and useful outcome of the project.

The finding of relatively high reliability of location measures, even on these very abstractly defined dimensions and with rather poorly defined research projects, was expected but gratifying. Location measurement is a matter for expertise, and these judges were experts in the field. Both the ease with which a true zero point for utility was defined, and its precision, were surprising. One technical reason for this success was that six obviously unattractive proposed research projects were carried through

the analysis, rather than being eliminated in the prescreening of proposed projects. The presence of these on the final list helped considerably in locating the true zero point. Happily, all six fell below it. A second and less interesting reason for the precision of the zero point may well have been that only one respondent was asked to make that particular set of judgments.

The difficulties at the decision stage resulted, of course, directly from the failure to define the decision options clearly enough and early enough. That's one mistake I believe I have learned not to make again.

3.3 Example 3: Indices of Water Quality

The work summarized in this example was performed by Dr. Michael F. O'Connor as his Ph.D. thesis (1972).

In 1968, the U.S. National Sanitation Foundation (known as NSF, but not to be confused with the National Science Foundation) published an index based on an additive combination of measures of nine parameters of water quality. The judgments were collected from more than 70 water quality experts. However, the index did not distinguish among possible uses of water and so left unanswered the question of whether different indices might be appropriate for different purposes. O'Connor set out to answer that question by developing two different indices. One described the quality of a surface body of water, treated as necessary, to be used as a public water supply. The other described the quality of a surface body of untreated water from the point of view of its ability to sustain a fish and wildlife population. These two uses will be abbreviated PWS (public water supply) and FAWL (fish

and wildlife) respectively. O'Connor's approach was to develop multi-attribute utility models for each use and then to examine the relationship between these models. At least moderate correlations were inevitable, but absence of very high correlations would indicate that at least two indices of water quality were needed.

3.3.1 Procedure - Eight experts on water quality located all over the country were the subjects. Four were university professors; others were officials in organizations responsible for water supplies.

Initially, 36 parameters of water were selected. In a mailed questionnaire, the experts were asked to rate the importance of each parameter for each of the two uses on a 1-to-100 scale by assigning 100 to the most important parameter and rating others relative to that parameter. (A variant on my proposed procedure, this one has the advantage that experts usually agree better on what is most important than on what is least important; but it also has the disadvantage of making it more difficult to preserve the ratio properties of the weight estimates.)

In a follow-up visit, each expert selected a subset (twelve or so) of the original 36 parameters and re-rated the importances of those he had selected. He also drew a function relating the relevant physical parameter continuum (e.g., pH) to quality; the function was required to have its maximum at 100 and its minimum at 0.

On the basis of the results of this visit, a second questionnaire, feeding back the results from other experts and asking for a re-rating of importances, was sent

out. It was followed by a second visit. For the second visit, the list of parameters was reduced to 17 for PWS and 11 for FAWL, in part by deletion of parameters considered by still other experts to be redundant with some that were retained. The main goal of the second visit was to achieve consensus on both importances and functions relating parameters to quality. The main tool used for this purpose was displays of all judgments obtained from questionnaire 2 and of average weights and functions. No expert objected to the parameter deletions; indeed, during the second visit four more parameters were deleted from the PWS list and two from the FAWL list. Table 3 shows the final parameters and normalized average importance weights. Most of the judges were willing to accept the average functions relating each physical parameter to quality as adequately representative of their own opinions but were much less willing to accept the average weights. The final functions, relating water quality to physical parameters and averaged over experts, were also accepted by most experts.

A final procedure consisted of preparing a number of imaginary water samples, described by parameter values on the relevant dimensions. Each expert was told the parameters of the sample, the scaled values of PWS and FAWL developed from the averaged data, those obtained from the expert's own weights combined with the average curves, and how the sample would score on the previously developed NSF index of water quality. Experts were invited to inspect these indices for the same use were very highly correlated; the lowest correlation between an average index and one prepared from an individual expert's judgments was .922, and when he changed some judgments the correlation rose to .956. Inter-correlations among PWS, FAWL, and the earlier NSF

PWS PARAMETER	NORMALIZED WEIGHTS	FAWL PARAMETER	NORMALIZED WEIGHTS
FECAL COLIFORMS	.171	DISSOLVED OXYGEN	.206
PHENOLS	.104	TEMPERATURE	.169
DISSOLVED SOLIDS	.084	pH	.142
pH	.079	PHENOLS	.099
FLUORIDES	.079	TURBIDITY	.088
HARDNESS	.077	AMMONIA	.084
NITRATES	.070	DISSOLVED SOLIDS	.074
CHLORIDES	.060	NITRATES	.074
ALKALINITY	.058	PHOSPHATES	.064
TURBIDITY	.058		
DISSOLVED OXYGEN	.056		
COLOR	.054		
SULFATES	.050		

TABLE 2: FINAL PARAMETERS CHOSEN FOR INCLUSION IN THE PWS AND FAWL INDICES

index were moderate, generally in the range from .6 to .8. Clearly, use does make a difference; a single water quality index is not good enough.

Linear approximations to the average curves were tried and generally produced very high correlations (e.g., .968) with the indices based on the average curves. An exception arose for certain water samples (chosen for realism) and the FAWL index, where the linear approximation produced correlations in the .70 region with the nonlinear index. This exception resulted from bad fits between the nonlinear function and its linear approximation for phosphates, turbidity, and dissolved solids, all of which were highly viable in the realistic water samples.

3.3.2 Comments - Most of the rather forceful methods used to obtain agreement in this study were necessary because of shortage of time with each expert and lack of opportunity for face-to-face discussion among the experts. While this procedure is not well designed to make experts feel happy with the final outcome, it did produce PWS and FAWL indices that seem serviceable for most purposes and that are clearly different. Face-to-face procedures would probably have produced very similar results but would have left the experts feeling happier about the indices finally developed.

O'Connor had considerable difficulty in getting his experts to understand the importance weighting method be used. It is unclear whether the difficulties were caused by shortage of time to explain and practice, or by the method itself; I suspect both.

Experts and O'Connor himself had difficulties with the additive model. One difficulty had to do with

toxic substances, such as pesticides. Both indices were made conditional on absence on these substances; their inclusion in even rather small concentrations would have made the water of unacceptably low quality, in the opinion of these respondents.

The other difficulty is more instructive. Both pH and fecal coliforms were important for PWS, but fecal coliforms were more than twice as important as pH. But low pH values (i.e., acid water) will kill the fecal coliforms and so may actually increase water quality. This relationship, so far, is clearly an instance of environmental correlation, not of violation of the underlying additive value model. However, a pH as low as 3.0 produces a water so unsatisfactory as an input to PWS that its quality is zero regardless of its merits on the other dimensions. Consequently, at this low pH level, the additive value model is violated. O'Connor handled this problem by using the additive model above 3.0 pH, and defining any water with pH of 3.0 or lower to have quality 0 for PWS. This definition produces an ugly discontinuity in the model but is otherwise unimportant since a pH of 3.0 or anywhere near it is rare indeed in water being considered as input to PWS.

4.0 CONCLUSION

This paper has reviewed three attempts in more-or-less applied settings to use multi-attribute utility measurement with a number of expert respondents. Three very different approaches to the problem of interpersonal disagreement are illustrated by the three examples. All seem to work. Comparing them, I feel that the procedure that used face-to-face discussion most heavily (the OCD example) was most successful in producing agreement; procedures depending on written or verbal feedback of other experts' judgments were clearly less so.

All three examples underline, in my view, the importance of simplicity in elicitation procedures. Amounts of respondent time ranged from a minimum of six hours to a maximum of two days per respondent in these examples; that is simply too short a time to teach any expert how to make sophisticated judgments about preferences among imaginary bets, and then collect a useful set of judgments from him, especially if a great deal of that time is taken up, as it should be, with discussion between him and other experts about the substantive issues lying behind the judgments.

So important does this issue of simplicity seem to me that our next major study will examine the following question: How well can a multi-attribute utility measurement procedure do by using an additive model, linear single-dimension utility functions for monotonic dimensions, and importance weights of 1, 0, and -1 only? The literature on unit weighting in multiple regression [Dawes and Corrigan (1974); Einhorn and Hogarth (1975)] suggests that unit weighting may work

surprisingly well, as does the literature on combining subtests (Wilks, 1938). I expect that high negative environmental correlations among dimensions of value can make such an approximation too simple. Still, if such an approximation is not too bad, what an enormous simplification of elicitation methods it offers us!

REFERENCES

Arrow, K.J. Social choice and individual values. New York: Wiley, 1951.

Bauer, V., & Wegener, M. Simulation, evaluation, and conflict analysis in urban planning. Proceedings of the Institute of Electrical and Electronic Engineers, 1975, 63, No. 3.

Dawes, R.M., & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 97-106.

Edwards, W. Social utilities. The Engineering Economist, Summer Symposium Series, 1971, 6, 119-129.

_____, & Guttentag, M. Experiments and evaluations: A re-examination. In Bennett, C. & Lumsdaine, A. (eds.), Experiments and Evaluations. New York: Academic Press, 1975.

_____, Guttentag, M., & Snapper, K. Effective evaluation: A decision theoretic approach. In Guttentag, M. (ed.), Handbook of Evaluation Research. Beverly Hills, CA: Sage Publications, 1975.

Einhorn, H.J., & Hogarth, R.M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Gardiner, P.C. The application of decision technology and Monte Carlo simulation to multiple objective public policy decision making: A case study in California coastal zone management. Unpublished Ph.D. thesis, University of Southern California, 1974.

_____, & Edwards, W. Public values: Multiattribute utility measurement for social decision making. Research Report 75-5, Social Science Research Institute, University of Southern California, 1975. Also in Schwartz, S., & Kaplan, M. (eds.), Human Judgment and Decision Processes. New York: Academic Press, 1975.

Guttentag, M., & Snapper, K. Plans, evaluations, and decisions. Evaluation, 1974, 2, 58-74.

Keeney, R.L. Utility functions for multi-attributed consequences. Management Science, 1972, 18, 276-287.

_____. Multiplicative utility functions, Operations Research, 1974, 22, 22-34.

O'Connor, M.F. The application of multi-attribute scaling procedures to the development of indices of water quality. Unpublished Ph.D. thesis, University of Michigan, 1972. (See also, O'Connor, M.F. The application of multi-attribute scaling procedures to the development of Indices of water quality. Report 7339, Center for Mathematical Studies in Business and Economics, University of Chicago, 1973.)

Raiffa, H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Mass.: Addison-Wesley, 1968.

_____. Preferences for multi-attributed alternatives. Memorandum RM04968-DOT/RC, The RAND Corp., 1969.

Torgerson, D. Venice: Everything is changing, middle-income hippies moving in where poor are moving out. Los Angeles Times, November 18, 1973.

Wilks, S.S. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 1938, 3, 23-40.

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)
Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce MacDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)
International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin
Office of Naval Research
Scientific Liaison Group
American Embassy – Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory
Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)
Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command (Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command (Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center (Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research Laboratory
U.S. Army Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 How to Use Multi-Attribute Utility Measurement for Social Decision Making	9	5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER [redacted] TR-76-3-1-T
7. AUTHOR(s) 10 Ward/Edwards		8. CONTRACT OR GRANT NUMBER(s) Prime Contract No: N00014-76-C-0074 Subcontract No: 75-030-0711
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Southern California Social Science Research Institute Los Angeles, CA 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209	11	12. REPORT DATE Aug[redacted]976
		13. NUMBER OF PAGES 67
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Decisions and Designs, Inc. Suite 600, 8400 Westpark Drive McLean, Virginia 22101 (under contract from the Office of Naval Research)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

12 69p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

15 N00014-76-C-0074, ARPA Order-3052

18. SUPPLEMENTARY NOTES

14 SSRI-TR-76-3-1-T

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multi-attribute utility	minimum plausible value
weighted linear average	maximum plausible value
utility functions	value independence
importance weights	environmental correlation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The thrust of this paper is that a public value is a value assigned to an outcome by a public, usually by means of some public institution that does the evaluating. This amounts to treating "a public" as a sort of organism whose values can be elicited by some appropriate adaptation of the methods already in use to elicit individual values. From this point of view, the interest of the problem lies in finding the appropriate adaptation of those methods, an adaptation that will take into account individual

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

392066

LB

disagreements about values, individual differences in relevant expertise, existing social structures for making public decisions, and problems of feasibility.

Multi-attribute utility measurement can spell out explicitly what the values of each participant (decision-maker, expert, pressure group, government, etc.) are, show how much they differ, and in the process can frequently reduce the extent of such differences. The exploitation of this technology permits regulatory or administrative agencies and other public decision-making organizations to shift their attention from specific actions to the values these actions serve and the decision-making mechanisms that implement these values.

The paper is structured around three examples. One is land use management; the specific example will be a study aimed at the decision problems of the California Coastal Commission. The decision-making body in this case is a regulatory agency exposed to a wide variety of social pressures from those with stakes in its actions.

The second example is concerned with administrative decision-making, specifically, with the process that the Office of Child Development of the U.S. Department of Health, Education, and Welfare used to develop its research program for the 1974 fiscal year.

The third example is more abstract; it concerns an attempt to develop a consensus among disagreeing experts on water quality, about a measure of the merits of various water sources for two purposes: the input, before treatment, to a public water supply, and an environment for fish and wildlife.